

To members of the TEI-C Board and Council
From Martin Mueller, chair, TEI-C Board

August 4, 2011

Dear colleague,

What follows is a long letter to you containing reflections and suggestions about goals and directions for the TEI-C in the years to come. This is not a set of formal proposals but a starting point for discussion. If the letter helps to frame that discussion it will have served its purpose.

Since I have never held office in the TEI-C and am in many ways a newcomer and outsider, I would like to turn that handicap into an opportunity and speak my mind. If some of what I say turns out to be just silly, it will be easy to ignore. If some of it is helpful, there may be a use for it for a conversation with our members --and more particularly the research libraries that are our main source of support--about what we can or should do for and together with them.

This is a rambling piece in which I think through various issues relating to the TEI that strike me as important and that at some point before the Würzburg meeting I want to raise with the membership in a more formal document. I hope that your responses will help separate the wheat from the chaff, not to speak of matters that I have overlooked. Questions that I return to, perhaps not in the best order concern

1. the ways in which the TEI has succeeded and failed,
2. the lack of tools that help scholarly end users make use of the added value that is in principle created by structural encoding,
3. whether the TEI should have a seat at the table of "Big Data,"
4. what role it can play in "research data lifecycle management,"
5. what role it might play in collaborative transcription of manuscripts by amateur scholars

6. how to address its hybrid role as a standards body, a scholarly society, and a membership association
7. how to depend less on the support of a few institutions and move towards more broadly based support of more members with a lower fee structure
8. how to establish a proper balance between financial contributions from North America and Europe
9. whether the TEI-C is a top-heavy organization relative to the size of its membership and the scope of its operations

I do not think that this letter contains any confidential data, and you may share it as you see fit. I want to stress, however, that in this letter I speak entirely for myself, talking about things as I see them after listening to you and talking with you, as well as with various outsiders, about TEI matters.

TEI as an enabling technology for the world of letters

At the heart of the TEI is a markup language designed to model in a digital environment texts that originated in a world of print and manuscripts. The TEI operates in a world of "letters," an old-fashioned but useful word, and it is an enabling technology for text-centric research in history, linguistics, literature, philosophy, religion, and cognate disciplines. TEI encoding can certainly be used for other purposes, but unless it is accepted as a critical or important enabling technology by historians, linguists, literary scholars, philosophers, theologians, etc., there is not much point in having it in the first place.

TEI markup uses XML, which is a business technology. There is nothing wrong with that. The Greek alphabet was adapted around 800 BCE from a Semitic alphabet, almost certainly because Phoenician and Greek traders wanted to keep track of what they were doing. But the earliest surviving piece of Greek writing is a hexametric line on a vase that proclaims to be Nestor's drinking cup and speaks of the aphrodisiac powers it bestows on its owner. I believe that T. S. Eliot typed the *Wasteland* on

a machine that at the time was still a relatively new piece of business technology. Commerce and culture or "Besitz und Bildung" are very old and close cousins.

TEI is a scholarly technology, and its "value proposition" --an ugly but useful term from the world of business-- consists in the promise for deeper or more extensive inquiries that properly encoded TEI texts offer to text-centric scholars. By comparison HTML is not a scholarly technology, although it certainly can be used for scholarly purposes and has been used so with varying success.

Because the TEI is a dialect of XML it is bound by two fundamental assumptions of that language. The first of these is the assumption that you can neatly separate the structure of a document from the manner of its display. The second is that every text is an "ordered hierarchy of content objects" in the sense that a single hierarchical model accounts for all structural aspects of a text. Most text-centric scholars will have problems with these guiding assumptions. On the other hand, both of them work often enough to be heuristically useful. In Wallace Stevens' poem "Connoisseurs of Chaos" there is the wonderful line "The squirming facts exceed the squamous mind." It is a good line to remember when dealing with just about any attempt to contain human language, whether written or spoken, within strict rules.

Success and failure

From one perspective, the TEI has exceeded expectations. Virtually all digital editions of primary texts with any claim to scholarly standards use it. TEI is the lingua franca of digital scholarly editing on a global basis. You find it in editions of Buddhist sutras, New Zealand and Pacific island texts, Greek inscriptions, French manuscripts of the *Roman de la rose*, the Hengwrt manuscript of the *Canterbury Tales*, slave narratives of the American South, or the historical records of the State Department. TEI has been used in all the large-scale university- library-based digitization projects of primary texts at Indiana, Michigan, North Carolina,

Virginia, and the Library of Congress. The same is true of European encoding projects.

That is the success story. But now consider a thought experiment where you ask the chairs of history, literature, linguistics, philosophy, and religion departments of the world's 100 top universities to write a sentence or short paragraph about the TEI. These would be very short sentences or paragraphs. The one message you would not get from them is the recognition that the TEI offers an important enabling technology for work in their disciplines. Results would be a little better (but not much) if you asked the chief librarians of those universities or their technology officers. You would get a little more knowledge but no ringing endorsement of the TEI as a key technology for some disciplines. You'd get the eye-rolling and shoulder-shrugging that comes with humoring a bright but wayward child. At least that has been my experience. In a forthcoming piece for the Profession issue of the MLA Jerry McGann observes that TEI stands for "terra incognita." The pun may be cruel, but I fear there is something to it.

Decoding the encoded

What accounts for this disconnect and what could or should we do about it? The TEI is about "encoding." The point of any encoding is decoding at the other end. A book is an encoding for human readers who decode it. What about the decoding of TEI encoded documents? There is a value and pleasure (masochistic at times) for TEI mavens who wrangle recalcitrant textual phenomena into the corset of a TEI schema. There may be an even deeper pleasure in cursing the inadequacies of the out-of-the-box schema and customizing it to a point at which it expresses all or most of the "squirming things" that do not fit into the "squamous mind." The pleasure of "difficulté vaincue." I googled the term because of my shaky grasp of French orthography. At the top of the list comes "Satisfaction, Difficulté Vaincue" showing a rock climber looking down rather smugly. What would we do without Flickr and Google? (<http://www.flickr.com/photos/45169745@N00/46485696/>)

But what about the added value of TEI specific encoding for the historian, linguist, philosopher, literary critic etc.? How can they decode or get at it, and what does it do for them? The answer is that for the most part they cannot get at it at all. I remember a conversation with a librarian who said something like "Oh yes, those TEI texts. We put them through the Lucene indexer and that's pretty much it." In principle, TEI encoding increases the query potential of the digital surrogate that is created by it. In practice, most of that query potential is ignored by the indexing and search software through which the encoded texts are mediated. Or if it is not ignored, it is used as instructions for XSLT style sheets to render the XML in HTML. As a result, the scholarly end users who encounter TEI-encoded texts almost never encounter them in an environment where they can take advantage of the distinct affordances of that encoding.

You can spend a lot of time explaining to your colleagues in an English department that it is a wonderful thing for texts to be encoded in TEI because it offers a much more robust, granular, and flexible way of storing textual data in digital form. But if what they see is browser-rendered HTML and if what they search might as well be a plain text file, it is not easy to persuade them of the value of this robust, granular, and flexible encoding. It does nothing to help them with their current project. Thus it is not much of an exaggeration to say that for ordinary scholarly users, TEI encoded texts right now offer no advantage over plain-text, html, or epub texts. Nietzsche once exclaimed in exasperation: "Was hilft mir der echte Text wenn ich ihn nicht verstehe?" or "What use is the true text if I don't understand it?" One might vary this into "What use is the encoded text if I cannot decode it?"

This is of course not true of the many and mostly small-scale projects that offer carefully and sometimes exquisitely curated editions of this or that set of documents. I yield to nobody in my admiration for the philological and technical ingenuity or beauty of design that has gone into

these projects. But allow me to be quite blunt for a moment and say something that I have believed strongly ever since I learned about the TEI in the mid-nineties. While small-scale and lovingly curated projects will have an important place in the ecosystem of digital textuality, the TEI will become an increasingly marginal niche player unless it stakes out what scholars and librarians recognize as a compelling role in the curation and exploration of the large-scale corpora that are changing the ways in which research is done in many disciplines.

Big Data and Research Data Lifecycle Management

Oren Sreebny at the University of Chicago wrote a running blog about a recent Princeton workshop about "research data lifecycle management" (<http://blog.orenblog.org/>). In reading through it, I was struck by two quotations attributed to Brian Athey, a bioinformatics researcher from Michigan, in a talk about Big Data 2011:

It's difficult to incentivize researchers to share data.

Agile data integration is an engine that drives discovery.

"Interoperability" is the word that negotiates the space between these two statements. This is an area where TEI practitioners have done less in the past than they might have and need to do a lot more in the future. The ability to scan very rapidly across large masses of textual data is a quite recent phenomenon. It has been dominated in the popular and the scholarly mind by "Googling" or shallow forms of text retrieval that work across billions or trillions of words in uncurated and often messy data. In the lifetime of the TEI, Natural Language Processing tools and methods have made enormous strides. Thus 1991 the British linguist John Sinclair wrote in the preface to his *Corpus Concordance Collocation*:

Thirty years ago when this research started it was considered impossible to process texts of several million words in length. Twenty years ago it was considered marginally possible but lunatic. Ten years ago it was considered quite possible but still lunatic. Today it is very popular.

That was then. The British National Corpus, a stratified sample of 100 million words of contemporary English, was released in 1994. Over the past year, Mark Davies at Brigham Young University has released two 400 million word corpora of Historical American English (1800-) and Contemporary American English (1990-2010). The Text Creation Partnership, probably the most ambitious academic transcription enterprise, will by 2015 produce TEI-encoded digital versions of some 70,000 public domain texts of English texts before 1800, with a word count of somewhere between 5 and 10 billion words. Google's n-gram corpus provides limited access to 500 billion words in some five million books.

These changes of scale have attracted the attention of scholars and librarians in ways that the TEI has not. Ask a modal chief librarian about the TEI, and s/he hardly knows what to say. Ask her or him about Google Books or Hathi Trust, and s/he hardly knows how to stop talking. Then there are the NLP and information retrieval folks who love Big Data and argue that if you have very large data sets you will get enough 'signal' even if the noise level is high. That is partly true, but it is also a convenient fiction because whatever you may want to say about reducing noise level in large data sets, it is very boring work.

Data curation has emerged as an important cross-disciplinary topic in recent years. In a network-based environment, there has been a lot interest in user-generated curation. "Crowdsourcing" and "dispersed annotation" are terms that float in various contexts. The interest is due to a growing recognition that more data may not be enough. There are lots of things you may have to do "to" the data before you can do useful things "with" them. If you follow the rhetoric of the Hathi Trust it is still preoccupied with "more" and takes a Herodotean delight in pure enumeration. On its home page you learn that its 9,410, 319 total volumes stretch for 111 miles and weigh 7,653 tons. But behind the scenes there is more interest in data quality and steps that would increase the interoperability of large textual data sets.

(<http://www.hathitrust.org/home>)

Is there a role for the TEI in the "research data lifecycle management" of large primary source materials in the humanities? Not of everything, but of very large data sets that repay higher levels of curation because such curation enables inquiries that could not be done otherwise. Corpus linguists tend to be uninterested in or are openly skeptical about the structural annotation of text corpora and like to think that "smart" processing of "dumb" data can give them all they need. They may be right for their domains, but in the humanities it is unlikely that researchers will have the programming skills that are common in the NLP community. There is likely to be a considerable research potential in coarse but consistent structural annotation that is applied across very large data sets and made the basis for query tools that can support quite granular queries through the intersection of coarse criteria.

You may not want to do this for millions of books, but within a given language it may be worth doing for tens of thousands of texts in ways that start from some "seed corpus" and rely on user-driven and collaborative curation and augmentation. In recent correspondence, Neil Fraistat and Doug Reside came up with the acronym CRIPT for "curated repository of important primary texts" -- an entity that could live as a special collection or inner ring inside much larger aggregates like Hathi Trust or Google Books. But while such entities would be smaller by orders of magnitude than the aggregate of all digitized texts, each of them would be considerably larger than could be managed by individuals or departmental projects. We are in a world of changed scale with requirements for robust infrastructure and highly skilled technical staff that are most easily envisaged as part of large academic libraries or GenBank like institutes.

In such a vision of things the TEI can easily claim a role as a critical enabling technology for any structural modeling of document structure, and it is really the only game in town. It can also claim that pieces of such a

vision exist here and there, whether in the French MEET Project, the large-scale German corpora under development by the Berlin-Brandenburg Academy and the TextGrid project, or the Text Creation Partnership.

What is needed for such a vision to succeed is a collective interest in query tools, and a much stronger commitment to interoperability across diverse data sets. By and large, the TEI has followed a "perl" ethos of reveling in the fact that the same thing can be done in many ways. I remember a conversation over a decade ago with a young programmer, who told me that he liked Python better than perl. I had never heard of Python and asked him why. He said that in Python there was typically only one way of doing a particular task and that made it easier to write consistent and reusable code. Much later I came across the Zen of Python and especially liked #13

There should be one -- and preferably only one -- obvious way to do it.

There are also quite a few occasions of looking at TEI encoding when I have wistfully remembered "Flat is better than nested" (#5). And I have spent many hours listening to the tirades of very gifted programmers about the inconsistency and unnecessary complexity that make TEI texts difficult or impossible to process. Within the confines of an individual project you can typically work around such problems. But if you want to work across corpora the problems very quickly turn into roadblocks. It would help a lot if the TEI adopted a "Pythonic" ethos and encouraged its practice with various combinations of preaching and nagging.

The rise of xquery, a turn towards decoding, and the need for interoperability

Unnecessary divergence and inconsistency --eloquently and often lamented by Mark Olsen -- are only one reason for the fact that much of TEI encoding is "lost in translation." Much more important is the fact

that the development of XML aware search engines has lagged behind encoding. Charles Rosen, not a friend of original instrument performances, once wrote that sonatas like the *Appassionata* were written for an instrument that did not exist until thirty years after Beethoven's death, when Theodore Steinway, now in New York, added cross-stringing to the cast iron frame that Chickering had pioneered in Boston in the 1820's. In the XML world we do not yet have query tools that match the scale and complexity achieved by 30 years of tweaking SQL and 15 years of making SQL work with the Web. But over the past decade xquery and the emergence of XML databases like eXist or BaseX give considerable hope for the future. There is also the potential for mimicking some XML awareness by loading information about some elements as "positional attributes" into the indexes of Lucene or the CQP query language.

These are very promising developments, but if you look at encoding from the perspective of subsequent decoding by an end user who wants to constrain a result set by combining various elements you may discover that less is more. The crossing of even a handful of criteria creates a great variety of possible searches.

I conclude from these reflections that it might be a good idea for the TEI to concentrate for a while on the problems and opportunities of encoding and decoding "at scale." This would shift the focus of attention to a quite basic set of tags and to the ways in which they could be made to work across many data and create, so to speak, a TEI-API for query tools. The TextGrid concept of "Kernkodierung" or "base line encoding" points in that direction. The point is not to prevent elaboration or complexity where it is necessary, but build levels of complexity on a pyramidal base to which texts from different collections can be reduced and which sits considerably above the word token. Think of a "highest common factor" rather than the lowest common denominator.

By way of example, a TEI-P5 set of about 60 elements, not counting the

header, is all that is needed to encode the large variety of Early Modern English texts in the TCP. Making these 60 tags work better, promoting their consistent use, and explaining to a lay audience how they work and what you can do with them may for quite a while be a better use of TEI resource than refining or adding to the current element set. The three percent of "squirring facts" that wriggle out of any corset will always be of greater interest to humanists with their innate fondness for the *haecceitas* of things. But there is a greater pay-off in the boring work of designing and maintaining robust and consistent practices for the limited set of elements that can deal quite adequately with the 97% of cases that do not resist the squamous mind.

The issues I'm reflecting about here are hardly specific to the TEI. In an earlier posting to the TEI list I quoted from the Lexus project at the Max Planck Institute for Psycholinguistics:

Lexicography in general is a domain where uniformity and interoperability have never been the operative words: depending on the purpose and tools used different formats, structures and terminologies are being adopted, which makes cross lexica search, merging, linking and comparison an extremely difficult task. LEXUS is also an attempt at putting an end to this problem. Being based on the Lexical Markup Framework (LMF), an abstract model for the creation of customized lexicons defined following the recommendations of the ISO/TC 37/SC 4 group on the standardization of linguistic terminology, LEXUS allows on the one hand to create purpose-specific and tailor-made lexica, and on at same time assures their comparability and interoperability with other resources.

Whether they will succeed is far from clear. But if the TEI wants to engage the growing interest in corpus-based inquiry "at scale" it needs to put much more emphasis on achieving high levels of interoperability for a limited element set. In terms of a product cycle, we may need to focus more on implementation and execution of the "Top Sixty" tags than on invention or design of new things. And marketing the "Top Sixty" with flair and simpler documentation may pay off as well. In this context it

has been instructive to me to follow the correspondence between Google's Ranjith Unnikrishnan and several council members about transforming Google books into curatable P5 rough cuts. I draw from it the tentative conclusion that we lack a good short overview of the TEI. I have a somewhat similar response to the excellent TEI by Example site, where the documentation is written a little too much from within. Too much of TEI stuff is written from within.

To sum up this part of my letter, it may be a good idea for the TEI in the coming years to focus on scholarly end users not as encoders of data but as decoders of already existing TEI texts. This involves finding answers to two questions:

1. How can we make sure that the benefits of encoded texts are in fact delivered to the scholarly users for whom these encodings were intended in the first place?
2. Should there be a place for the TEI at the Big Data table, and if so, what should we do to get a place at that table?

Agribusiness and organic farming

While I believe that the TEI's engagement with Big Data is important I don't want to be heard as saying it should be the only game in town. From the scholarly end user's perspective the critical problem is how to decode the encoded or unlock the query potential with which encoders have enriched the source data. This critical problem applies equally at the micro- and macro-levels, and it may be that good solutions will come from scaling up small or mid-scale projects. A few months ago, Katherine Rowe, the chair of English at Bryn Mawr, who has a deep interest in digital tools as a way of promoting research opportunities for undergraduates, drew my attention to the work of Robert Binkley, an American historian of the 1930's and leader in the WPA local history project. In 1935 Binkley wrote a remarkable essay in the Yale Review called "New Tools for Men of Letters."

(<http://www.wallandbinkley.com/rcb/articles/newtools-output.html>)
A lead sentence in that essay reads: " The new graphic arts devices are, I believe, capable of working the other way — as implements for a more decentralized and less professionalized culture, a culture of local literature and amateur scholarship."

I wrote to her:

I looked at Robert Binkley's essay and was as taken with it as you have been. It's a virtual contemporary of Walter Benjamin's essay on the work of art in a mechanical age, and in its way it worries about not unrelated things. The technologies have changed, but the ideological framing has stayed very much the same. A lot of humanists (and not only humanists) like to shop at farmers' markets. I think of Google and Hathi Trust as textual agribusiness. Success for the digital humanities will have to consist of connecting the agribusiness of it with local farming and organic gardening.

To which she replied: "Lovely! That's a blog entry if I ever saw one... "
I quote from this exchange at such length because it may be a useful way of gesturing towards a space of institutional culture and feeling that is good to keep in mind while planning for the future.

Helping scholarly end users to decode the encoded

Back to the business of helping scholarly end users to decode the encoded. I could imagine the TEI taking a much more active and formal interest in tool development that focuses on getting out what encoders put in. There is actually a lot of tool building going on. How all the activities are connected remains an open question. As I understand it, King's College is about to release a document publishing system that abstracts a common template from their various projects. Excellent work with eXist has been done at Brown and Victoria (Canada). Nebraska is dipping several toes into eXist. The BaseX people have expressed an interest in the TEI. At the US State Department Joe Wicentowski has been building a site that publishes the Foreign Relations of the United States with a

conceptually simple model that relies solely on eXist and xquery. Julia Flanders and Scott Hamlin have recently got NEH grants that fall broadly into the category of outreach and evangelizing. Then there is Text-Grid, which does a lot but could use a friendlier interface.

I am not saying that the TEI should or could go directly into the business of tool building. I could, however, envisage a meta-role as shaper of a continuing, active, and structured conversation about goals and principles of good design, always focusing on the question: "How can scholarly end users get out the stuff that encoders put in?" I can see circumstances in which the Mellon Foundation could be interested in funding such a conversation through workshops of one kind or another. In a more speculative mode (in a cognitive and economic sense) I could envisage the TEI becoming a consultant and broker that would help member institutions hire each other's programmers for specific projects. Angie's List for TEI.

You might say that what I am arguing for is already happening at sufficient scale and that there is no virtue in organizing it further. That is a serious argument, but I have a hunch that there are quite a few dots that need connecting and that the TEI can play a useful role in connecting them.

Can the TEI play a role in the collaborative curation of manuscripts?

Returning for a moment to encoding and specifically to encoding of manuscripts, there seems to be a lot of interest in the collaborative transcription of all manner of manuscript materials, whether medieval stuff, Venetian state documents, or Civil War letters. In some of these projects palaeographical niceties are critical, while in others you want a simple diplomatic transcription that produces machine actionable texts. I don't know a whole about this, and I also don't know how much of a role the TEI plays in such simple and pragmatic projects. The discussions of manuscript matters on the TEI list tend towards what Michael Witmore calls "the philologically exquisite." It would be a good thing for the TEI if

encoders at all levels, from the humble to the most learned, would think of the TEI as the default choice for manuscript transcription. Are we the first name that comes to the minds of folks in local and state historical societies when they have a transcription project? Can we lower the entry barriers to TEI based manuscript encoding?

The hybrid role of the TEI as a standards body, scholarly society and membership association

I now turn to financial and institutional questions. The TEI is an odd body, a hybrid of a standards body, a scholarly society, and an institutional membership association. As a standards body it belongs in the world of acronyms like EAD, METS, DocBook, etc. At least that is where you find it when you want to start a new file in oXygen and confront your choices of a document type. A standards body has a single purpose: to develop, maintain, and improve a particular standard. A scholarly society is more like a chat room. It has no single purpose or product but provides a space of communication, which ranges from gossip to scholarly papers. Its outcomes are meetings, proceedings, and journals. A membership organization is more like a lobby: its members justify their support in terms of particular goals that they want to happen or keep from happening.

The TEI-C is a delicate and somewhat awkward balancing act between those three things. Let me describe the stress points in this balance in terms of the threats to each of the three identities of the TEI-C. To begin with the TEI-C as a scholarly society, there is not much of a threat because there is very little to threaten. I base my observation on financial records in a Quicken database, which may not record all transaction accurately, but is probably not seriously misleading. Between 2005 and 2010 there have been in any year between 17 and 25 paid-up individual subscribers. There is not much continuity and there is no trend line. Most of the members of the Board or Council have never been individual subscribers (myself included), and hardly any have been steady subscribers over the years.

There is a perfectly innocent explanation for this. Many of the Board and Council members come from universities that pay institutional membership fees, and the hardworking members of both bodies may feel that their in-kind contributions of time and energy are more than enough. I completely accept this argument, but it is worth pointing out that the sizable and quite real international “TEI community” does not appear to feel a need to express its sense of belonging through subscription to a society. From a financial perspective, the revenue from individual subscriptions is a trivial part of the budget. I am agnostic on the question whether a membership drive would be worth the effort. I am equally agnostic on the question whether teaming up with the ADHO would make a substantial difference from a financial perspective.

From a financial perspective all the Digital Humanities societies are shoestring operations when compared with other scholarly societies. ASECS, the society for 18th century studies, has a budget of about \$300,000, and much of its income comes from individual subscriptions. The Renaissance Society of America is much richer. Its annual budget is about a million dollars, and it has non-trivial endowment funds.

As a standards body the TEI does not face a threat from any competitor. If there is a threat it comes from people who believe that structural text encoding simply is not worth the effort. This is a widely shared view in the NLP and information retrieval community, which is an important source of text analysis tools and protocols. As I have said above, the TEI needs to hold its own in that argument and cannot afford to take the utility of its encoding standard for granted.

Membership association, research libraries, and the TEI budget

That brings us to the TEI as a membership organization, and this is a good point to produce some budget figures. By the end of this year the TEI is likely to have a cash balance of not quite \$150,000, which is a good testimony to a prudent board and its officers. *Grosso modo* income

and expenditures have been around \$100K+/-10K. We are, I believe, rather better off than other DH organizations, but we also face a threat to the main source of our income because our major supporters, large research libraries are likely to take an increasingly hard look at the value proposition of their substantial membership fees.

In 2010, the TEI-C collected \$92,000 in membership fees from 70 member institutions. \$60,000 or two thirds of that revenue came from twelve institutions paying membership fees of \$5,000. Nine of them are American research libraries, three of them are institutions in Canada, England, and France. In the current year, two American libraries have canceled their memberships, and a third appears likely to do so. In France, we have lost a partner but gained a new one. We will lose the Canadian partner next year, but may gain a partner in Germany. If you look at the number of \$5,000 memberships between 2005 and 2012, there is, alas, a trend line in the numbers 14, 12, 15, 12, 9, 9.

Conversations with some librarians have confirmed my sense of what is going on. There was a time not so long ago when a chief librarian would readily authorize a membership of \$5,000 if the cause was generally worthy and some faculty member took an interest in it even if the library was not especially engaged. \$5,000 may have been chump change then, but it is real money now and likely to stay that way. No librarian will authorize an expenditure of that size unless it can be justified in terms of specific institutional priorities.

I had an interesting conversation about this with Merrilee Proffit, whom I met a decade ago at the Pisa meeting when she was part of the medieval manuscript group. She is now a senior program officer of the OCLC Research Library Partnership, which is the successor of the RLG. She told me that the RLG used to charge membership fees of \$25,000 but has now gone to a sliding scale between \$1,000.00 and \$10,000. From this I gather that the TEI-C needs to move towards a more distributed mem-

bership model with more members and lower fees. But even with lower fees the “worthy cause” argument is likely to fall on deaf ears.

Asymmetric funding between North America and Europe

There is another asymmetry in the current funding stream, as is apparent from this table, which lists the number of memberships in each country, the revenue from each country, and the revenue as a percentage of total membership income:

country	members	total	percentage
US	38	\$58,750.00	63.8
Canada	9	\$8,500.00	9.2
France	6	\$7,441.00	8.1
UK	15	\$7,425.00	8.1
Germany	7	\$2,468.24	2.7
Ireland	2	\$1,750.00	1.9
Norway	5	\$1,435.00	1.6
Belgium	1	\$500.00	0.5
Denmark	1	\$500.00	0.5
Netherlands	2	\$500.00	0.5
New Zealand	1	\$500.00	0.5
Taiwan	1	\$500.00	0.5
Slovenia	1	\$479.00	0.5
Austria	1	\$473.00	0.5
Czech Rep.	2	\$334.00	0.4
Bulgaria	2	\$266.00	0.3
Hungary	1	\$250.00	0.3

2010 is a typical year in terms of the geographical distribution of funding. In looking at this table it is hard not to feel that the TEI-C would be

better off if roughly half of its income came from North America and the other half from Europe.

In thinking about how to raise more money from European countries one must realize that the American way of raising money through memberships does not work very well in Europe. There is more “top-slicing” of funds, and there are no categories in the budgets of library or institute directors that memberships can legitimately be charged to. That was made very clear to me in recent conversations with Gerhard Lauer and Fotis Jannidis in Göttingen. It was confirmed by Merrilee Proffit who said that her organization very much wanted to straddle the Atlantic and that the Europeans wanted to do this too but it was difficult to find a funding model. This seems like a silly problem, but it appears to be genuine.

On the revenue side, then, the TEI-C could move along quite comfortably at its current rate of operations with a budget of around \$100K, where \$50K comes from North America and its equivalent in euros and pounds comes from Europe. The North American funds would be raised through a more distributed model of more memberships at lower cost, while the European funds would be raised in some other fashion. It is not, after all, a whole lot of money.

What about spending?

On the spending side there are a number of uncertainties, and I am not sure I understand them fully. Until and through 2010, much of the technical and administrative work of the TEI-C was done at four host institutions that provided both cash and in-kind services but also were paid for additional services. This system was phased out in last year’s reorganization. There are no anchor institutions anymore. Instead the TEI-C has a looser system of “partners” (in principle unlimited in number) who contribute cash and in-kind services but receive no payment. The budget will commit funds for paid services, but these may change

from year to year and are not in advance tied to any individual or institution.

In the current year, if I read the budget correctly, nobody gets paid for anything, but some accounting services will be farmed out. How this will work out in practice remains to be seen.

Is the TEI-C too top-heavy?

I am now going to say things that people may not like, but I am going to say them anyhow. That is the outsider's privilege, and if you recognize my remarks as based on incorrigible ignorance you can laugh me out of court. Coming from the outside, looking at the organization charts of various roughly comparable organizations, and considering the scope of operations and size of membership, I am inclined to think that the TEI-C is a top-heavy organization. I am reluctant to say this, particularly in view of the fact that there was a substantial reorganization last year. But how do you explain to prospective members why an organization with some 70 institutional memberships requires a council of twelve and a board of eight elected members, not counting non-voting Board members, whether representatives of partner institutions or appointed officers like myself?

How can an organization with 70 members justify an overhead of two dozen people who by virtue of their office can lay a claim for reimbursement for their travel expenses, which add up very quickly and are in fact the largest item in our budget? I would have a very hard time explaining this mismatch to my chief librarian if I tried to persuade her to commit to a full membership. So I think we need to give some serious thought to making the organization leaner. There are different ways of doing this without damaging the work. One could in fact make a plausible argument that a leaner structure would be nimbler and more effective. Below I sketch a fairly radical model that would, however, be compatible with the articles of incorporation. Think of it as a straw man

proposal to shoot at while thinking about the problems that made me suggest it in the first place:

1. The current board and council are replaced by a board of 12 directors, at least seven of them elected and the others elected or appointed as the membership sees fit.
2. The board of directors divides into a technical committee of seven members and a general committee of five members.
3. The chair of the general committee is also the chair of the board.
4. All offices in the organization are held by current members of the board.
5. The entire board meets once a year at the members' meeting. The technical committee meets separately on at least one other occasion.

There would be an expectation that directors of the TEI-C would have their travel expenses borne by their home institutions wholly or in part. This "pay for honor" model has much going for it and works equally well on both sides of the Atlantic. I was told by Fotis Jannidis that this would be much easier to fund in Europe than a membership model. It also is psychologically more attractive. Instead of paying a membership fee to an abstract organization, a department chair or library head pays for a colleague's expenses as a form of recognition or career development. There will certainly be cases where a potential director comes from an institution that simply cannot pay very much, but there are ready ways of dealing with this.

Would a technical committee of seven do a worse job than a committee of twelve? Not necessarily. Much of the intellectual work of the TEI happens on and off the general listserv. People cycle on and off the council, but in their off-years they are still engaged in the intellectual life of the community. Even now the work of the council consists less of generating new proposals than of reviewing, filtering, and editing proposals that come in from the field. Seven people are not necessarily worse at

this than a dozen and may in fact do better. Besides, the five other directors are likely to have strong and useful views about technical matters. So I would not see this straw man proposal as involving a reduction in the intellectual capital of the current council.

You can see that I am a “less is more” person except when it comes to writing long memos. Whether or not this straw man has real legs to stand, he should make us ask whether we are a little top-heavy and need to think about ways of becoming leaner, not simply as a matter of cutting costs but also in terms of becoming more effective.

Dr. Johnson said of *Paradise Lost* that nobody wished it to be any longer. You may feel the same about this letter. So I stop here and hope that my rambling thoughts will at least spur a useful discussion.

Sincerely

Martin Mueller